

Source and Accuracy Statement for the Report
Computer Use in the United States: 1997

SOURCE OF DATA

The data for this report come from the October 1997 Current Population Survey (CPS). This month's survey uses two sets of questions, the basic CPS and the supplement. The Bureau of the Census conducts the basic CPS every month and asks supplementary questions during certain months.

Basic CPS. The basic CPS collects primarily labor force data about the civilian noninstitutional population. Interviewers ask questions concerning labor force participation about each member 15 years old and over in every sample household.

October 1997 supplement. In addition to the basic CPS questions, interviewers asked supplementary questions on internet and computer use.

Sample Design. The CPS sample includes coverage in all 50 states and the District of Columbia. The Census Bureau continually updates the sample to account for new residential construction. The Census Bureau divides the United States into 2,007 geographic areas. In most states, a geographic area consists of a county or several contiguous counties. In some areas of New England and Hawaii, the Census Bureau uses minor civil divisions instead of counties. We select a total of 754 geographic areas for sample. About 50,000 occupied households are eligible for interview every month. Field representatives are unable to obtain interviews at about 3,200 of these units. This occurs when the occupants are not found at home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Bureau of the Census has redesigned the CPS sample several times. These redesigns have improved the quality and accuracy of the data and have satisfied changing data needs. The Census Bureau completely implemented the most recent changes in July 1995.

Estimation procedure. This survey's estimation procedure adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by age, sex, race, Hispanic/non-Hispanic origin, and state of residence. This adjustment is called the post-stratification ratio estimate. The independent estimates are based on:

- The 1990 Decennial Census of Population and Housing.
- An adjustment for undercoverage in the 1990 census.
- Statistics on births, deaths, immigration, and emigration.
- Statistics on the size of the armed forces.

The independent population estimates include some, but not all, undocumented immigrants.

ACCURACY OF THE ESTIMATES

Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. A sample survey estimate has two possible types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error, but the full extent of the nonsampling error is unknown. Consequently, one should be particularly careful when interpreting results based on a relatively small number of cases or on small differences between estimates. The standard errors for CPS estimates primarily indicate the magnitude of sampling error. They also partially measure the effect of some nonsampling errors in responses and enumeration, but do not measure systematic biases in the data. (Bias is the average over all possible samples of the differences between the sample estimates and the desired value.)

Nonsampling variability. We can attribute nonsampling errors to several sources including the following:

- Inability to obtain information about all cases in the sample.
- Definitional difficulties.
- Differences in the interpretation of questions.
- Respondents' inability or unwillingness to provide correct information.
- Respondents' inability to recall information.
- Errors made in data collection such as recording and coding the data.
- Errors made in processing the data.
- Errors made in estimating values for missing data.
- Failure to represent all units with the sample (undercoverage).

For the October 1997 basic CPS, the nonresponse rate was 6.3% and for the supplement the nonresponse rate was an additional 4.7% for a total supplement nonresponse rate of 10.7%.

CPS undercoverage results from missed housing units and missed people within sample households. Compared to the level of the 1990 Decennial Census, overall CPS undercoverage is about 8 percent. Undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites. The post-stratification ratio estimate described previously partially corrects for bias due to undercoverage. However, biases exist in the estimates to the extent that missed people in missed households or missed people in interviewed households have different characteristics from those of interviewed people in the same age-sex-race-origin-state group.

A common measure of survey coverage is the coverage ratio, the estimated population before the post-stratification ratio estimate divided by the independent population control. Table A shows CPS coverage ratios for age-sex-race groups for a typical month. The CPS coverage ratios can exhibit some variability from month to month, but these are a typical set of coverage ratios.

Table A. CPS Coverage Ratios

Age	Non-Black		Black		All People		Total
	M	F	M	F	M	F	
0-14	0.929	0.964	0.850	0.838	0.916	0.943	0.929
15	0.933	0.895	0.763	0.824	0.905	0.883	0.895
16-19	0.881	0.891	0.711	0.802	0.855	0.877	0.866
20-29	0.847	0.897	0.660	0.811	0.823	0.884	0.854
30-39	0.904	0.931	0.680	0.845	0.877	0.920	0.899
40-49	0.928	0.966	0.816	0.911	0.917	0.959	0.938
50-59	0.953	0.974	0.896	0.927	0.948	0.969	0.959
60-64	0.961	0.941	0.954	0.953	0.960	0.942	0.950
65-69	0.919	0.972	0.982	0.984	0.924	0.973	0.951
70+	0.993	1.004	0.996	0.979	0.993	1.002	0.998
15+	0.914	0.945	0.767	0.874	0.898	0.927	0.918
0+	0.918	0.949	0.793	0.864	0.902	0.931	0.921

For additional information on nonsampling error including the possible impact on CPS data when known, refer to Statistical Policy Working Paper 3, An Error Profile: Employment as Measured by the Current Population Survey, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978 and Technical Paper 40, The Current Population Survey: Design and Methodology, Bureau of the Census, U.S. Department of Commerce.

Comparability of data. Data obtained from the CPS and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Use caution when comparing results from different sources.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The Bureau of Labor Statistics redesigned questionnaire to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The Census Bureau also modified the supplemental questions for adaptation to computer-assisted interviewing, but did not change definitions and concepts. Because of these and other changes, one should use caution when comparing estimates from data collected in 1994 and later years with estimates from earlier years.

Data users should also use caution when comparing estimates from this report (which reflects 1990 census-based population controls) with estimates for 1993 and earlier years (which reflect 1980 census-based population controls). This change in population controls had relatively little impact on summary measures such as means, medians, and percentage distributions. It did have a

significant impact on levels. For example, 1990 based population controls caused about a 1-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1994 and later years will differ from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain subpopulation groups than for the total population.

Since no independent population control totals for people of Hispanic origin were used before 1985, compare Hispanic estimates over time cautiously.

For more information on the introduction of the new questionnaire, the modernized data collection methods, and the introduction of new population controls based on the 1990 census, see “Revisions in the Current Population Survey Effective January 1994” in the February 1994 issue of *Employment and Earnings* published by the Bureau of Labor Statistics.

Note when using small estimates. Because of the large standard errors involved, summary measures (such as medians and percent distributions) probably do not reveal useful information when computed on a base smaller than 75,000. Take care in the interpretation of small differences. For instance, even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Sampling variability. Sampling variability is variation that occurred by chance because a sample was surveyed rather than the entire population. Standard errors, as calculated below, are primarily measures of sampling variability, but they may include some nonsampling error.

Standard errors and their use. The Census Bureau had to make a number of approximations to derive, at a moderate cost, standard errors applicable to estimates from this report. Instead of providing an individual standard error for each estimate, we have provided two parameters, a and b, to calculate standard errors for each type of characteristic. These parameters are in Tables B and C.

The sample estimate and its standard error enable one to construct a confidence interval, a range that would include the average result of all possible samples with a known probability. For example, if all possible samples were surveyed under essentially the same general conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples. However, one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Data users may also use standard errors to perform hypothesis testing. This is a procedure for distinguishing between population parameters using sample estimates. One common type of hypothesis is that two population parameters are different. An example of this would be comparing the number of men who were part-time workers with the number of women who were part-time workers.

One can perform tests at various levels of significance. A significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. To conclude that two parameters are different at the 0.10 level of significance, for example, the absolute value of the estimated difference between characteristics must be greater than or equal to 1.645 times the standard error of the difference.

The Census Bureau uses 90-percent confidence intervals and 0.10 levels of significance to determine statistical validity. Consult standard statistical textbooks for alternative criteria.

For information on calculating standard errors for labor force data from the CPS which involve quarterly or yearly averages, changes in consecutive quarterly or yearly averages, consecutive month-to-month changes in estimates, and consecutive year-to-year changes in monthly estimates; see “Explanatory Notes and Estimates of Error: Household Data” in the corresponding *Employment and Earnings* published by the Bureau of Labor Statistics.

Standard errors of estimated numbers. One can obtain the approximate standard error, s_x , of an estimated number from this report by using the formula

$$s_x = \sqrt{ax^2 + bx} \quad (1)$$

Here x is the size of the estimate and a and b are the parameters in Tables B or C associated with the particular type of characteristic. When calculating standard errors from cross-tabulations involving different characteristics, use the set of parameters for the characteristic which will give the largest standard error.

Illustration

In October 1997 there were 3,267,000 unemployed men in the civilian labor force. Use the appropriate parameters from Table B and Formula 1 to get

Number, x	3,267,000
a parameter	-0.000018
b parameter	2,957
Standard error	97,300
90% conf. int.	3,106,900 to 3,427,100

The standard error is calculated as

$$s_x = \sqrt{-0.000018 \times 3,267,000^2 + 2,957 \times 3,267,000} = 97,300$$

The 90- percent confidence interval is calculated as $3,267,000 \pm 1.645 \times 97,300$.

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Standard errors of estimated percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on both the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the parameter from Table B or C indicated by the numerator.

One can obtain the approximate standard error, $s_{x,p}$, of an estimated percentage by using the formula

$$s_{x,p} = \sqrt{(b/x)p(100 - p)} \quad (2)$$

Here x is the total number of people, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \leq p \leq 100$), and b is the parameter in Table B or C associated with the characteristic in the numerator of the percentage.

Illustration

In October 1997 there were 102,158,000 households, 36.6 percent had a computer in the household. Use the appropriate parameter from Table C and Formula 2 to get

Percentage, p	36.6
Base, x	102,158,000
b parameter	2,068
Standard error	0.22
90% conf. int.	36.2 to 37.0

The standard error is calculated as

$$s_{x,p} = \sqrt{(2,068/102,158,000) \times 36.6 \times (100 - 36.6)} = 0.22$$

The 90-percent confidence interval of the percentage of households with computers is calculated as $36.6 \pm 1.645 \times 0.22$.

Standard error of a difference. The standard error of the difference between two sample estimates is approximately equal to

$$s_{x - y} = \sqrt{s_x^2 + s_y^2} \quad (3)$$

where s_x and s_y are the standard errors of the estimates, x and y . The estimates can be numbers, percentages, ratios, etc. This will represent the actual standard error quite accurately for the difference between estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration

In October 1997 there were 2,572,000 unemployed men 20 years of age or older and 2,409,000 unemployed women 20 years of age or older. Use the appropriate parameters from Table B and Formulas 1 and 3 to get

	x	y	difference
Number	2,572,000	2,409,000	163,000
a parameter	-.000018	-.000018	-
b parameter	2,957	2,957	-
Standard error	86,500	83,800	120,400
90% conf. int.	2,429,700 to 2,714,300	2,271,100 to 2,546,900	-35,100 to 361,100

The standard error of the difference is calculated as

$$s_{x - y} = \sqrt{86,500^2 + 83,800^2} = 120,400$$

The 90-percent confidence interval around the difference is calculated as $163,000 \pm 1.645 \times 120,400$. Since this interval includes zero, we can not conclude with 90 percent

confidence that the number of unemployed men is greater than the number of unemployed women.

Accuracy of state estimates. The redesign of the CPS following the 1980 census provided an opportunity to increase efficiency and accuracy of state data. All strata are now defined within state boundaries. The sample is allocated among the states to produce state and national estimates with the required accuracy while keeping total sample size to a minimum. Improved accuracy of state data has been achieved with about the same sample size as in the 1970 design.

Since the CPS is designed to produce both state and national estimates, the proportion of the total population sampled and the sampling rates differ among the states. In general, the smaller the population of the state the larger the sampling proportion. For example, in Vermont approximately 1 in every 400 households was sampled each month. In New York the sample was about 1 in every 2,000 households. Nevertheless, the size of the sample in New York is four times larger than in Vermont because New York has a larger population.

Computation of standard errors for state estimates. Standard errors for a state may be obtained by computing national standard errors, using formulas described earlier, and multiplying these by the appropriate factor, f , from Table D. An alternative method for computing standard errors for a state is to multiply the a and b parameters in Table B or C by f^2 and then use these adjusted parameters in the standard error formulas.

Illustration

In October 1997 there were 7,011,000 households in New York, 32.4 percent of which had a computer. Use the appropriate parameter from Table C and Formula 2 to get

Percentage, p	32.4
Base, x	7,011,000
b parameter	2,068
Standard error	0.80
Factor, f	0.94
New York standard error	0.75

Thus, the standard error on the estimate of the percentage of households in New York state with a computer is approximately $0.75 = 0.94 \times 0.80$.

To obtain state parameters, multiply the parameters in Table C by f^2 in Table D for the state of interest. The value of f^2 for New York is 0.89. Thus, for Total or White household characteristics, such as computer ownership, in New York this gives
 $a = -.000012 \times 0.89 = -0.000011$ and $b = 2,068 \times 0.89 = 1,841$.

Computation of a factor for groups of states. The factor adjusting standard errors for a group of states may be obtained by computing a weighted sum of the squared factors for the individual

states in the group and taking the square root of the result. Depending on the combination of states, the resulting figure can be an overestimate.

The squared factor for a group of n states is given by

$$f^2 = \frac{\sum_{i=1}^n \text{POP}_i \times f_i^2}{\sum_{i=1}^n \text{POP}_i}$$

where POP_i is the state population and f_i^2 is obtained from Table D. The 1998 civilian noninstitutionalized population from the CPS for each state is also given in Table D.

Illustration

Suppose a factor for the state group Illinois-Indiana-Michigan was required. The appropriate squared factor would be

$$f^2 = \frac{8,925,000 \times 0.99 + 4,467,000 \times 1.37 + 7,288,000 \times 0.92}{8,925,000 + 4,467,000 + 7,288,000} = 1.05$$

Multiply the a and b parameters by f^2 , 1.05, to obtain parameters for the state group. Alternatively, calculate standard errors with unadjusted parameters and multiply the result by f , 1.02, to get standard errors for this state group.

Table B. Parameters for Computation of Standard Errors for Labor Force Characteristics
October 1997

Characteristic	a	b
Labor Force and Not In Labor Force Data Other than Agricultural Employment and Unemployment		
Total ¹	-0.000018	2,985
Men ¹	-0.000033	2,764
Women	-0.000030	2,530
Both sexes, 16 to 19 years	-0.000172	2,545
White ¹	-0.000020	2,985
Men	-0.000037	2,767
Women	-0.000034	2,527
Both sexes, 16 to 19 years	-0.000204	2,550
Black	-0.000125	3,139
Men	-0.000302	2,931
Women	-0.000183	2,637
Both sexes, 16 to 19 years	-0.001295	2,949
Hispanic origin ²	-0.000206	3,896
Not In Labor Force (use only for Total, Total Men, and White)		
	+0.000006	829
Agricultural Employment		
Total or White	+0.000782	3,049
Men	+0.000858	2,825
Women or		
Both sexes, 16 to 19 years	-0.000025	2,582
Black	-0.000135	3,155
Hispanic origin ²		
Total or Women	+0.011857	2,895
Men or		
Both sexes, 16 to 19 years	+0.015736	1,703
Unemployment		
Total or White	-0.000018	2,957
Black	-0.000212	3,150
Hispanic origin ²	-0.000102	3,576

¹ For not in labor force characteristics, use the Not In Labor Force parameters.

² Hispanics may be of any race.

Table C. Parameters for Computation of Standard Errors for Computer Literacy Estimates October 1997

	Total or White		Black		Hispanic ¹		API	
Characteristic	a	b	a	b	a	b	a	b
PERSONS								
Educational Attainment	- 0.000011	2,369	- 0.000109	2,680	- 0.000148	3,052	- 0.000246	2,164
Persons by Family Income	- 0.000026	4,901	- 0.000260	5,611	- 0.000556	9,456	- 0.000638	5,611
Income	- 0.000012	2,454	- 0.000120	2,810	- 0.000249	4,736	- 0.000327	2,810
Marital Status, Household & Family Characteristics	- 0.000019	5,211	- 0.000221	7,486	- 0.000443	12,616	- 0.000627	7,486
Poverty	- 0.000039	10,380	- 0.000307	10,380	- 0.000617	17,493	- 0.000869	10,380
FAMILIES, HOUSEHOLDS, OR UNRELATED INDIVIDUALS								
Income	- 0.000013	2,241	- 0.000119	2,447	- 0.000354	4,124	- 0.000352	2,447
Marital Status, Household & Family, Educational Attainment, Population by Age or Sex	- 0.000012	2,068	- 0.000077	1,871	- 0.000261	3,153	- 0.000279	1,871
Poverty	+0.000102	2,442	+0.000102	2,442	+0.000172	4,115	+0.000102	2,442

¹ Hispanics may be of any race.

Table D. Factors for State Standard Errors and Parameters and State Populations: 1997

State	f	f ²	Population
Alabama	1.00	1.01	3,307,000
Alaska	0.39	0.15	432,000
Arizona	0.98	0.96	3,468,000
Arkansas	0.77	0.59	1,921,000
California	1.13	1.27	23,969,000
Colorado	0.96	0.93	2,953,000
Connecticut	1.00	1.00	2,520,000
Delaware	0.47	0.22	565,000
District of Columbia	0.41	0.16	423,000
Florida	0.99	0.97	11,304,000
Georgia	1.18	1.40	5,620,000
Hawaii	0.60	0.36	870,000
Idaho	0.51	0.26	895,000
Illinois	1.00	0.99	8,925,000
Indiana	1.17	1.37	4,467,000
Iowa	0.84	0.71	2,184,000
Kansas	0.80	0.64	1,922,000
Kentucky	0.96	0.92	3,009,000
Louisiana	0.97	0.94	3,231,000
Maine	0.60	0.36	974,000
Maryland	1.17	1.38	3,885,000
Massachusetts	0.90	0.81	4,742,000
Michigan	0.96	0.92	7,288,000
Minnesota	1.05	1.11	3,523,000
Mississippi	0.80	0.64	2,041,000
Missouri	1.17	1.37	4,060,000
Montana	0.44	0.20	679,000
Nebraska	0.65	0.42	1,241,000
Nevada	0.66	0.44	1,259,000
New Hampshire	0.62	0.38	897,000
New Jersey	0.90	0.82	6,150,000
New Mexico	0.63	0.40	1,285,000
New York	0.94	0.89	14,002,000
North Carolina	0.97	0.94	5,609,000
North Dakota	0.40	0.16	480,000
Ohio	1.01	1.02	8,548,000
Oklahoma	0.84	0.71	2,487,000
Oregon	0.93	0.86	2,518,000
Pennsylvania	0.98	0.95	9,288,000
Rhode Island	0.55	0.30	754,000
South Carolina	1.00	1.01	2,852,000
South Dakota	0.41	0.17	543,000
Tennessee	1.16	1.34	4,152,000
Texas	1.10	1.21	14,313,000
Utah	0.65	0.43	1,429,000
Vermont	0.42	0.18	456,000
Virginia	1.21	1.47	5,078,000
Washington	1.22	1.49	4,243,000
West Virginia	0.62	0.38	1,454,000
Wisconsin	1.09	1.19	3,934,000
Wyoming	0.34	0.12	366,000

NOTE: For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.